



Library Data Science Committee

Framework Recommendations

University Libraries
The University of North Carolina at Chapel Hill
March 2020

Committee Members

Nandita Mani
Michelle Cawley
Lorin Bruckner
Jason Casden
Adam Dodd
Amanda Henley
Matt Jansen
Jamie McGarty
Morgan McKeehan
Sarah Morris
Therese Triumph
Jessica Venlet
Joe Williams

Contents

Executive Summary	3
Recommendations	6
Recommendation 1: Communication & Branding	6
R1.A: Communication & Branding	7
Recommendation 2: Assessment & Reporting	8
R2.A. Assessment	8
R2.B. Reporting	10
Recommendation 3: Human Resources & Employee Engagement	10
R3.A Reskilling	10
R3.B Performance Management and Professional Development	13
R3.C New Hires	14
Recommendation 4: Library Data Science Priorities and Partnerships	18
R4.A Develop Partnerships	18
R4.B Communication with Partners	23
R4.C Joint or Co-Funded Positions	24
Recommendation 5: Creating and Expanding Services	24
R5.A Services to Expand	24
R5.A New Services to Create	25
Recommendation 6: Space & Infrastructure	27
R6.A Library Space and DS @ Carolina	27
R6.B Technical Infrastructure	27
Appendix A: Executive Summary of Institutional Interviews	29
Appendix B: Interview Questions for Library Counterparts at Exemplar Institutions	31
Appendix C: Skills Matrix Survey	32
Appendix D: Survey for UNC Partners and Stakeholders	37
Appendix E: Stakeholder Matrix	41
Appendix F: Environmental Scan	43
Executive Summary	43
Methods	43
Results	44

Executive Summary

The University Libraries at the University of North Carolina at Chapel Hill has created a framework to expand data science support of students, faculty and researchers. This framework is centered on a set of recommendations put forward by the Library Data Science Committee. The Committee's recommendations are based on multiple data sources, including an environmental scan (see [Appendix F](#)), interviews with library counterparts at exemplar institutions (see [Appendix A](#) for high level summary and [Appendix B](#) for interview questions), preliminary results from a library staff Skills Matrix Survey, and Committee experience and expertise. The Library Data Science Committee identified six categories of recommendations around:

- Communication & Branding
- Assessment & Reporting
- Human Resources & Employee Engagement
- Cultivating Partnerships
- Expanding and Creating Services
- Space & Infrastructure

Table 1 below provides a summary of the high-level goals and specific recommendations within each category. Additional details around each recommendation category follow.

To seize the opportunity to engage around Data Science @ Carolina initiative (hereafter DS @ Carolina) at the outset, the University Libraries should **establish an Implementation Team**. Any delays in engaging with campus partners and other stakeholders will set us back in terms of being seen as a partner and engaging in high level collaborations. To meet the needs of DS @ Carolina and to engage as full partners, we envision that the Implementation Team will:

- Develop mechanisms to evaluate needs of DS @ Carolina stakeholders on an ongoing basis.
- Oversee reskilling of University Libraries' staff to meet DS @ Carolina needs.
- Create position descriptions and engage in the hiring process for recruitment so University Libraries is well-positioned to support DS @ Carolina.
- Engage in periodic, regular evaluation of University Libraries' skills and services around data science.
- Establish and evaluate strategic objectives and priorities for University Libraries support of DS @ Carolina.
- Address organizational and cultural barriers to meeting the anticipated needs around DS @ Carolina.
- Provide ongoing assessment of tools, technologies, software, and services that are necessary to support DS @ Carolina.

If established, the Implementation Team should **apply change management principles** to prepare and support current staff for changes around services and partnerships that will be required of our organization to support DS @ Carolina. Specifically, a structured approach should be sought to provide clear and concise messaging around how the University Libraries supports DS @ Carolina, to minimize staff turnover, assure high morale, and align performance goals with this strategic initiative.

Table 1: Summary of Recommendations and Goals

Recommendation 1: Communication & Branding	
Goals	Includes Recommendations Around:
Showcase our expertise and capacity.	<ul style="list-style-type: none">Library Web SiteSocial MediaDonor & Campus EngagementBranding and Promotional MaterialsLibrary Newsletter & Library MeetingsCommunication via Data Science Labs
Cultivate data science partnerships.	
Provide clarity for external partners and patrons.	
Develop relationships with donors.	
Recruit new talent to our organization.	
Recommendation 2: Assessment & Reporting	
Goals	Includes Recommendations Around:
Showcase our expertise and capacity.	<ul style="list-style-type: none">Assessment and Evaluation PlanUNC Data Science NeedsData Science Spaces Assessment TeamData Science Skills AssessmentAnnual Overview
Cultivate data science partnerships.	
Develop relationships with donors.	
Recommendation 3: Human Resources & Employee Engagement	
Goals	Includes Recommendations Around:
Bridge the skills gap within University Libraries to support data science activities on campus.	<ul style="list-style-type: none">ReskillingPerformance Management and Professional DevelopmentNew Hires
Incentivize library staff to gain skills relevant to data science.	
Provide ample professional development opportunities.	
Develop a tiered approach to providing services around data science.	
Recommendation 4: Library Partnerships & Data Science Priorities	
Goals	Includes Recommendations Around:
Cultivate data science partnerships around research and curriculum integration.	<ul style="list-style-type: none">Library Priorities and PartnershipsConcierge RolesMechanisms for CollaborationFormal Communications ChannelsJoint-Funded Positions with Campus Partners
Transform existing services into an immersive model where library staff are integrated into research projects.	

Recommendation 5: Creating and Expanding Services

Goals	Includes Recommendations Around:
Meet Carolina's needs around data science.	<ul style="list-style-type: none">• Services to Expand• Services to Create
Cultivate data science partnerships.	

Recommendation 6: Space & Infrastructure




Goals	Includes Recommendations Around:
Cultivate community, interdisciplinarity, and catalyze new partnerships around data science.	<ul style="list-style-type: none">• Data Science Labs• Formal Space Use Policy• Technical Infrastructure Review Team• Infrastructure Partnerships• Technical Consultation Services Program• Use Policy for Libraries' technical infrastructure• Business Plan to Manage Infrastructure Costs
Establish a group to review the Libraries' current technical infrastructure.	
Create space use policies.	

Recommendations

This section lays out recommendations around six areas, including:

- Communication & Branding
- Assessment & Reporting
- Human Resources & Employee Engagement
- Cultivating Partnerships
- Expanding and Creating Services
- Space & Infrastructure

Recommendations are categorized as 1, 2, or 3 (or a combination of these) which refers to the following timelines:

-  1 Within 1 year
-  2 1 to 2 years
-  3 Consider after DS @ Carolina is established

Recommendation 1: Communication & Branding

An intentional strategy around branding and communication for internal and external purposes will allow us to showcase our expertise and impact to stakeholders. This will assist with building relationships with donors and campus partners, recruiting new staff, and providing clarity around our services and spaces to serve DS @ Carolina.

More than 60 institutions were reviewed for the environmental scan. Based on this review, we found that few libraries provide clear communications and branding around their data science support and partnerships. This is true even among institutions with strong data science programs.

Communication and branding should be done in collaboration with Library Communications and Development and should occur through multiple channels including the Library website, social media, signage, internal meetings, and events for campus partners and other stakeholders including the public (e.g., workshops, trainings, etc.). The individual(s) serving in a concierge role will have a critical role with respect to internal and external communication (see [Library Data Science Concierge](#)). Details on specific recommendations are provided below. Regardless of how the campus initiative moves forward, the Libraries should consider, in partnership with Library Communications, how to address communication and branding around data science and other services.

Communication & Branding Goals

- Showcase our expertise and capacity.
- Cultivate data science partnerships.
- Provide clarity for external partners and patrons.
- Develop relationships with donors.
- Recruit new talent to our organization.

R1.A: Communication & Branding

The Committee proposes the following recommendations around communication and branding:

1

Library Web Site: Establish a One Library approach to illustrate our support of data science efforts in terms of services, staff, spaces, events and workshops, and through our partnership with the DS @ Carolina Initiative. Those navigating our website should be able to clearly identify University Libraries' as a partner in DS @ Carolina and identify individuals working in data science related areas. Upon establishment of the campus-wide DS @ Carolina Initiative, ensure University Libraries is clearly indicated as a campus partner in their marketing materials and web pages.

1

Social Media: Create a strategy to ensure our work in data science and as partners to DS @ Carolina are highlighted in a One Library fashion via Facebook, Twitter, and other media outlets as needed.

1

Donor & Campus Engagement: Engage our donors and future prospects through Windows Magazine and the Gazette as a way of sharing how University Libraries is engaged in data science efforts at Carolina including, but not limited to, DS @ Carolina. Include stories of impact such as those where University Libraries has:

- Partnered with researchers and faculty on data science projects.
- Assisted students in acquiring data skills.
- Engaged with faculty on curriculum integration.
- Contributed to the Carolina Next (or DS @ Carolina) strategic framework.

1

Create a Brand. Create a graphic identity and tagline that reflects our values and promise to our constituents around serving campus data science needs. As part of this effort, create a messaging strategy to keep internal and external audiences informed and abreast of new developments at the Libraries related to DS @ Carolina.

Develop brochures, PowerPoint slide templates, and other promotional documents and/or marketing materials that staff can use to promote data science events and initiatives. Create signage for any data related physical spaces that are developed. Promote select data science research projects through a website, newsletter, and other media channels.

1

Library Newsletter & Library Meetings. Use internal communication opportunities to keep staff aware of data science initiatives and how the Library is engaged with DS @ Carolina. Highlight staff involved with data science projects, presentations, publications, and grants through our newsletter and at Department Heads and Library All Staff meetings. Invite data science partners (students, faculty, etc.) to present and/or co-present with Libraries staff to share their experiences around partnering with the Library.

2

Communication via Research Hubs and/or Data Science Spaces. Use our current spaces and, assuming University Libraries dedicates one or more spaces as data science labs on campus, use the labs to support communication and branding efforts as follows:

- Host data-related events for UNC affiliates and the public such as lectures, workshops, symposia, annual data day, and hackathons.

- Promote partner projects and share tools and technologies developed or used in collaboration between University Libraries and campus partners.

Recommendation 2: Assessment & Reporting

A clear assessment and evaluation program will ensure we are actively meeting the needs of our stakeholders. Assessment should include a plan for identifying resources needed for information, technology, and space. Evaluation plans should be created to monitor our progress with regards to the recommendations outlined. Reporting mechanisms should be created (see [Communication & Branding](#)) to highlight the progress and impact the University Libraries is making with DS @ Carolina. Details on specific recommendations are provided below.

Assessment & Reporting Goals

- Showcase our expertise and capacity.
- Cultivate data science partnerships.
- Develop relationships with donors.

R2.A. Assessment

2

Assessment and Evaluation: Establish a robust assessment and evaluation plan for each of the following core areas:

- Instruction and Curriculum Integration
- Research Support
- Priority Partnerships
- Library Spaces
- Library Computing, including infrastructure
- Library Skills and Services (e.g., data curation, data storage and management, classes and workshops)

Assessment and evaluation plans should identify specific metrics for each of these categories, to be measured on a regular basis (e.g., annually). For example, in the area of curriculum-integration, we can identify what topics we teach, how many sessions are course-integrated versus standalone workshops, how many sessions are taught, the number of students we reach, and if possible, for sessions in which we teach alongside faculty, we can include an evaluation component for library-based workshops/sessions that gauge acquisition of skills and knowledge.

As part of an assessment and evaluation plan we should establish performance indicators such as those detailed below.

Indicators of our success with research engagement may include:

- Number of grants that include Libraries staff for data science-related support.
- Number of high-profile projects where the Libraries is identified as a formal partner.
- Increase in quantity of collaborations around data science with priority partners on campus.
- Number of co-authored publications between Libraries staff and faculty/researchers where the Libraries' contributions pertain to data science.
- Number of Libraries staff with high level of expertise embedded in research projects.

- Number of projects with established project charters or memoranda of understanding (MOUs).
- Number of co-funded and joint positions.
- New positions or roles that show evidence of engagement with our partners.
- Acknowledgment of Libraries in data science articles, dissertations, and other publications.
- Opportunities for Libraries staff to provide substantive contributions in Data Science @ Carolina symposiums, conferences, or workshops.

Indicators of our success around curriculum integration and instructional support may include:

- Number of consultations (i.e., one-on-one or class/group) around data science.
 - Use of LibAnalytics should be reviewed to ensure statistics are entered consistently by all staff to ensure results are meaningful.
 - Tracking of online consultations should also be done to reflect support for distance students.
- Staff with a high level of expertise are embedded in research projects rather than conducting instruction and 1:1 consultations.
- Our stakeholders access asynchronous and reusable content around data science (e.g., recorded presentations and tutorials).
- Dual appointments for Libraries staff.

1 2

UNC Data Science Needs: The Implementation Team should assess campus needs around data science periodically through focus groups, surveys, or other methods. This should be done in the near term to prepare for supporting DS @ Carolina, soon after the program is established, and regularly thereafter. The Library Data Science Committee collated a list of campus stakeholders to include in future conversations around data science needs that may be used as a starting point.

1 2

Periodically Assess Skills Around Data Science. Use the skills matrix survey (see [Appendix C](#)) or a similar tool to evaluate data science skills across all University Libraries staff. A skills survey can be used to:

- Compare available expertise with campus needs to identify gaps that may inform new or modified positions.
- Identify staff with expertise who can engage in peer-to-peer training or other methods to reskill existing staff.
- Identify new staff as potential candidates for joint-funded positions.

2

Data Science Spaces Assessment Team: The Libraries Research Hubs provide several existing spaces for digital research that are well managed and used. Once new library services and campus partnerships are better understood, a Space Assessment Team can analyze changes needed for current or new spaces.

1

Evaluate Research Hubs & Potential for Data Science Spaces: Given that there will be significant overlap in services and staff between data science and what we currently offer via the Research Hub, and that we want to minimize confusion for patron and partners, an evaluation of all Hub

locations and potential for new data science spaces should be considered. The evaluation should consider overall data science needs and associated services regarding space, resource/staffing and whether the Research Hub needs to be re-envisioned and/or re-branded.

R2.B. Reporting

2

Annual Overview: As part of our annual report, provide a summary of data science efforts at the Libraries that is linked to University Libraries strategic priorities.

2

Internal Reporting Mechanisms: Use Department Heads and All Staff meetings to highlight the work of various people and/or teams related to DS @ Carolina. Encourage staff to conduct lightning rounds, share announcements, and invite partners to our meetings to share how they have collaborated with University Libraries around data science.

Recommendation 3: Human Resources & Employee Engagement

According to results from a Skills Survey administered to department heads, we have limited capacity in our organization for some of the data science skills needed to support anticipated needs of DS @ Carolina.

Workforce development should focus on the reskilling necessary to address the skills gap by increasing experience and exposure for existing staff (i.e., Figure 2: Tiers 0, 1, and 2).

To bridge gaps in expertise, new hires will be required. Expertise (i.e., Figure 2: Tier 3) requires a degree in a related field or significant time on task. In the near term, expertise can only come from new hires. University Libraries' staff with deep expertise around data science have limited ability to embed in research projects due to heavy commitments for instruction and consultations.

Human Resources Goals

- Bridge skills gap within University Libraries to support data science activities on campus through hiring and retraining.
- Incentivize library staff to gain skills relevant to data science needs on campus.
- Provide ample professional development opportunities.

Incentives to encourage building skills around data science include linking performance goals to the data science strategic initiative and providing time and funding for professional development opportunities. Reskilling will provide opportunities for career growth among current staff. One way to incentivize librarians with data science skills to partner with others on campus in teaching data science is to provide release time from their library duties, allowing them the time needed to teach as adjuncts. Details on specific recommendations are provided below.

R3.A Reskilling

Reskilling University Libraries' staff will be a critical element to meeting the demands of DS @ Carolina and will require ongoing attention. **University Libraries should explore opportunities to partner and engage with other campus units around reskilling.** The Libraries co-location with Odum Institute and strong relationship with the School of Information and Library Science (SILS) are major assets. Further, our relationship with SILS is unique based on more than 60 institutions reviewed as part of the environmental scan (see [Appendix F](#)).

1 2

Tiered Approach for Data Science-Based Professional Development. Library staff will be provided opportunities to expand skills relevant to supporting DS @ Carolina through participation in a semi-structured progression matrix (Figure 2). Skill development activities will be available for each tier, with staff progressing through tiers until they reach their chosen level of proficiency. Skills may be focused around particular content areas such as digital humanities as indicated below.

Figure 2: Proposed tiered service pyramid for delivering data science-related services and partnering around instruction and research.

Tier 3: Staff with high degree of expertise should:

- Partner on grants and research projects.
- Engage as adjunct professors.
- Perform original research.
- Provide high-level consultations.
- Provide limited support for instruction.
- Design curriculum-integrated instruction.

Tier 2: Staff with advanced technical skills should:

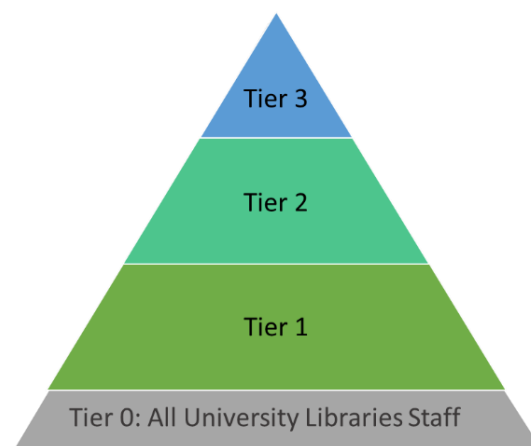
- Provide back-up on consultations as needed.
- Engage as adjunct professors.
- Partner on research projects.
- Discover and develop code to address internal Library needs.
- Deliver curriculum-integrated instruction.

Tier 1: Staff with introductory technical skills should:

- Provide introductory data-related consultations.
- Partner with Tier 2 staff to provide instruction.
- Attain introductory data literacy and technical skills (e.g., data manipulation in Excel).

Tier 0: All University Libraries Staff should:

- Triage data science-related inquiries.
- Understand that data science is a strategic priority and maintain an understanding of Library expertise and services around data science.



1

Tier 0: All staff will be provided opportunities for awareness raising through participation in introductory activities to learn basic vocabulary and how to triage data science-related requests. Opportunities may include:

- Internal lightning talks pertaining to data science staff projects.
- Selected webinars.
- Half-day internal training available to all University Libraries staff focused on jargon busting.

1

Tier 1: Subject Liaisons will be trained to provide first tier data science support for researchers in their areas of expertise, so they can provide data literacy instruction and data project support. They will refer requests for advanced data science support to data science librarians or partners. Opportunities for subject liaisons may include:

- Internal projects: Time-limited project to build particular skills (e.g., visualization, impact measurement, R, Python); report results to library staff via internal lightning talk.
- Short courses developed and offered in partnership with campus partners.
- Time on task with support from Tier 2 and 3 staff.
- Peer-to-peer training series.
- Selected courses via LinkedIn Learning.
- Partner with and support Tier 2 staff in providing instruction.
- Half-day internal training available to all University Libraries staff focused on introductory data skills in Excel.
- Campus-wide user groups (e.g., UNC Tableau Users Group, Excel Users of UNC Yammer Group).
- Libraries' Communities of Practice (e.g., Data Wrangling).

2

Tier 2: Staff with some experience will continue to develop proficiency through participation in internal Library and student research projects and by leading instruction. Opportunities may include:

- Peer-to-Peer training series.
- Time on task with Tier 3 staff support.
- Audit or enroll in courses within data science curriculum as part of work time, where appropriate.
- Short courses developed and offered in partnership with campus partners.
- Weekly block of time devoted to improving skills, learning new software, or working on research projects (e.g., digital humanities).
- Staff learning cohorts and research teams ranging from summer learning groups to research teams working on funded projects. For all projects, goals and plans should be developed with supervisor input and included in Annual Performance Plan documents for evaluation.

2

Tier 3: Staff with expertise will oversee and coordinate staff development through skills mentoring and involving Tier 2 staff in functional components of research projects. Opportunities to gain expertise may include:

- Audit or enroll in courses within data science curriculum as part of work time, where appropriate.
- Attend conferences or workshops.
- Present at conferences and workshops and/or co-present with subject matter experts.
- Joint appointments.

- Developing short courses in partnership with campus partners.
- Partner on grants and research projects.

R3.B Performance Management and Professional Development

Incentives are necessary to motivate staff to engage around data science. The recommendations below focus on performance management and professional development opportunities to increase employee engagement.

1

Align Annual Performance Goals. As staff attain foundational knowledge around data science and move up the tiered service pyramid (Figure 2), supervisors should work with staff to create goals that prioritize reskilling and ensure balance between current and future work. Specific recommendations include:

- Make real time changes to performance goals as necessary.
- Instruct supervisors on performance goals and assessment criteria for new duties related to data science.
- Rebalance workload for staff providing data science services to reflect Library priorities.

1

Professional Development. Provide opportunities for staff to attend and present at data science-related workshops and conferences. Specific recommendations include:

- Host additional Library Carpentry training events and encourage participation/certification among staff.
- Allocate time and funding needed to gain experience and knowledge acquisition around data science.
- Create and facilitate a library-wide Data Interest Group (DIG) to be held on a bi-monthly basis to share best practices and discuss current and potential opportunities.
- Partner with other campus groups to establish and participate in a campus-wide Data Interest Group (DIG).
- Provide staff who have new data science responsibilities additional professional development funds in the first 1-2 years of responsibilities. Priority should be given to those presenting on data science-related projects at regional or national conferences.

2

Annual Library Data Day: Host a regular data science & digital scholarship showcase to highlight staff contributions as well as program growth and overall impact. Consider including the opportunity for staff to work in teams on small scale projects and use Library Data Day to encourage cross-unit/cross-library participation and engagement.

R3.C New Hires



Data Science Positions: Over a two-year period, **create new library positions** across the library system to support data science efforts, considering the need for diverse educational backgrounds and experience (e.g., informatics, computer science, library & information science, data science). Positions should focus on digital humanities, humanities & social sciences, biomedical and the sciences. Compensation for current positions being re-formulated or redesigned to a more data-intensive role should be reviewed.

Tables 2 and 3 below lay out the Committee's recommendations around creating and hiring new positions in two phases. In addition to the full-time positions described below, the Committee recommends **recruiting at minimum 2 RA or GA positions** focused on data science to be situated across the University Libraries.

The explanations provided below may be used to draft final position descriptions, although adding more details around specific qualifications should be considered before posting announcements.

Table 2: Hiring Recommendations for Phase 1

Phase 1 Hiring		
Position	Explanation	Number of Positions
Type: Senior Administration Position Anticipated Rank: Full Librarian	This senior position would guide and coordinate the Libraries' Data Science initiative, direct the research agenda, foster new and existing partnerships, and communicate with the Library Leadership Team. Specific responsibilities for this position may include: <ul style="list-style-type: none"> • Guide research agenda and direct R&D. • Foster new and existing instructional and research partnerships with schools, divisions, institutes, centers, and industry partners. • Liaise with Library Communication and Web Development teams to address communication and branding related to library DS efforts. • Lead assessment and evaluation of Libraries DS engagement. • Communicate with Libraries leadership as needed. • Ensure coordination across staff and individual libraries. • Coordinate efforts around staff reskilling and creating data literacy competencies. 	1
Type: Data Science Librarians in areas of Social Sciences, Humanities, Natural Sciences, and Health Sciences	We recommend that resources are allocated to support developing our data science staff across the library system in Social Sciences & Humanities, Natural Sciences, and Health Sciences. Most researchers do not learn the computational tools and data management skills they will need to excel in today's data-driven research environment. Data Science Librarians	4

Table 2: Hiring Recommendations for Phase 1

Phase 1 Hiring		
<p>(Cluster Hire)</p> <p>Anticipated Rank: Up to Associate Librarian</p>	<p>provide advice and consultation to researchers on how to make the most of their data. Overall, the responsibility of a data science librarian involves data services such as assisting patrons with locating, acquiring, preparing, and managing data. Specific responsibilities for these positions may include:</p> <ul style="list-style-type: none"> • Discuss data problems with researchers, assess data needs, and design a suite of services based on user needs. • Develop and lead instructional sessions and outreach focused on management, interpretation, analysis, and visualization of varying forms of data. • Assist patrons in assessing repository options for data archiving and preservation. • Provide consultation and recommendations regarding the ethical issues of data use and reuse. • Develop and deliver data analysis services in response to current trends, campus needs, and Library priorities. • Use advanced skills with data cleaning/wrangling/normalization, regular expressions, web scraping, and APIs to support and collaborate with researchers on data-related research. • Identify, evaluate, and recommend new and emerging data science research tools and methods for the Libraries and UNC research community. • Continue to develop technical skills and proficiency with new data analysis tools and techniques. • Advise and recommend appropriate hardware and software to support data science work. • Develop and offer stand-alone workshops (in person and online) open to Libraries staff, students, faculty, and researchers. Instruction would focus on the use of data science methods and tools (e.g., R, Python, Tableau, geospatial applications in consultation with GIS Librarian). • Collaborate on research projects using quantitative and qualitative datasets. • Partner with researchers to identify and establish open science practices and policies, best practices in data management, and e-notebooks for lab environments. • Participate in data research projects across disciplines. 	
<p>Type: Data Science Instruction Librarian</p>	<p>This position (and associated position hired in Phase 2) would coordinate instruction related to data science, including data ethics. These positions would partner with liaisons to develop</p>	<p>1</p>

Table 2: Hiring Recommendations for Phase 1

Phase 1 Hiring		
Anticipated Rank: Up to Associate Librarian	<p>and deliver data literacy instruction, develop data literacy modules to integrate into courses taught by liaisons, and participate in efforts to coordinate reskilling of Libraries staff under direction of senior Libraries leadership. Specific responsibilities of this position may include:</p> <ul style="list-style-type: none"> • Coordinate instruction related to data science including data ethics. • Partner with liaisons to develop and deliver data literacy instruction. • Develop data literacy modules to integrate into courses taught by liaisons. • Participate in efforts to coordinate reskilling of Libraries staff under direction of senior Libraries leadership. • Advise and consult on pedagogy, instructional design, and curriculum-integrated instruction. • Work with liaisons to develop new course- and data-related instructional resources. 	
Phase 1: New Data Science Positions		6

Table 3: Hiring Recommendations for Phase 2

Phase 2 Hiring		
Position	Details	Number of Positions
Type: Humanities Data Science Librarian (R&D) Anticipated Rank: Up to Associate Librarian	<p>This position would focus on research and development of data sets that the Libraries can create and manage from our own collections, e.g., text mining data sets of oral history transcripts. Specific responsibilities for this position may include:</p> <ul style="list-style-type: none"> • Serve as technical lead on data science projects started by the Libraries. • Participate in Digital South program. • Meet the needs of library stakeholders by: <ul style="list-style-type: none"> ○ Preparing library metadata and collections to be machine-readable. ○ Creating collections (as data) from our digital collections and developing tools (e.g., APIs) for researchers to use to interface with those collections (e.g., NC Newspapers and transcripts from oral histories). • Spearhead research efforts focused on using AI to engage more deeply with existing collections. For example: Use 	1

Table 3: Hiring Recommendations for Phase 2

Phase 2 Hiring		
Position	Details	Number of Positions
	<p>OCR to identify (handwritten) names of the enslaved from documents in our collections.</p> <ul style="list-style-type: none"> Collaborate on the application of computational methods to assess and improve the management and discoverability of our digital collections, including extensive AV holdings. 	
<p>Type: Data Curation Librarian</p> <p>Anticipated Rank: Up to Associate Librarian</p>	<p>This is an existing gap in our organization and essential for supporting the Libraries' Preservation Pillar. Expanded data services and data archiving will require a dedicated specialist who can help bridge gaps in organizational infrastructure between the digital preservation stewardship committee, Library IT, and liaison librarians. This position will also contribute to technical infrastructure development and maintenance related to repository services. Specific responsibilities for this position may include:</p> <ul style="list-style-type: none"> Collaborate with Institutional Repository (IR) Librarian to assist UNC community in depositing datasets in the Carolina Digital Repository (CDR). Provide support for the IR program's maintenance and ongoing development of policies, procedures, and documentation relevant to dataset collections and open access. Contribute to aligning curation activities in collaboration with IR Librarian such as: facilitating dataset preparation, ingesting datasets to the repository, creating metadata, assessing rights and rights statements, understanding reuse issues and needs Align collecting and program development and facilitate curation activities with FAIR data principles. Collaborate on overall digital preservation activities with the Digital Preservation Stewardship Committee, Repository Services, University Archives, and Software Development. Collaborate with liaison and data science librarians around outreach for data archiving in the CDR or other data repositories, use of datasets available in the CDR, and other collections as data projects. Point person for connecting and networking with other data repositories and data curation archivists/librarians (such as Odum). 	1

Table 3: Hiring Recommendations for Phase 2

Phase 2 Hiring		
Position	Details	Number of Positions
Type: Data Science Instruction Librarian Anticipated Rank: Up to Assistant Librarian	This position would support the Data Science Instruction Librarian (hired in Phase 1). Specific responsibilities of this position may include: <ul style="list-style-type: none"> • Coordinate instruction related to data science including data ethics. • Partner with liaison librarians to connect data literacy to other core literacies. • Develop data literacy modules to integrate into courses taught by liaisons. • Participate in efforts to reskill Libraries staff. • Advise and consult on pedagogy, instructional design, and curriculum-integrated instruction. • Work with liaisons to develop new course- and data-related instructional resources. 	1
Phase 2: New Data Science Positions		3

Recommendation 4: Library Data Science Priorities and Partnerships

University Libraries has a variety of campus partners and should develop a strategic approach to identify and engage with priority partners around data science, including partnering with external groups such as industry, JSTOR, HathiTrust, etc.

Joint or co-funded positions will elevate the Libraries and will increase opportunities for partnership and high-level collaboration. Individuals in concierge roles will foster relationship building and communication. Details on specific recommendations are provided below.

Partnership Goals

- Cultivate data science partnerships around research and curriculum integration.
- Transform existing services into an immersive model where library staff are integrated into research projects.

R4.A Develop Partnerships



Library Priorities and Partnerships: The Implementation Team should develop a phased approach for engaging campus partners around data science in terms of curriculum integration and research. Priorities and partnerships should be reevaluated periodically. The table below lays out some potential opportunities for the Libraries to partner with other groups on campus.

Data Literacy Curriculum (Curriculum Integration): Build on our success in digital and information literacy to formalize a program that supports data literacy in the undergraduate curriculum.
Potential Outcomes
<ul style="list-style-type: none"> • Integrate the Libraries' course support and programming with the DS @ Carolina curriculum as it develops. • Create 'Ready-to-customize' data modules for data literacy in Ideas, Information, and Inquiry (Triple I) courses and others. • Data science workshops/short courses; working with partners to provide and organize tutors and/or mentors. • Provide workshops and curriculum modules drawing on our expertise in ethical use of data. • Develop a program that encourages students underrepresented in STEM disciplines to become data scientists. Establish summer workshops and short courses to support students interested in pursuing data science, particularly those students lacking prerequisites. • Collaborative workshops and seminars, possibly modeled on events such as Brown's DSCoV: Data Science, Computing, and Visualization workshops, and Academy in Context seminars.
Integration in Graduate Curriculum (Curriculum Integration): Work with campus partners to expand teaching collaborations. Provide RA, GA, or Carolina Academic Library Associate (CALA) position(s) devoted to data science work in the Libraries.
Potential Outcomes
<ul style="list-style-type: none"> • Expand current teaching collaborations, specifically with regards to key concepts of library science that overlap with data science (i.e., data description and visualization, workflow and reproducibility, and ethical problem solving). • Expand and make more structured our graduate research assistantships, potentially along the lines of the CALA program. If these (or a subset of these) were to focus on data science, we could train students to scale our work of supporting faculty with integrating data into the classroom. This could follow the model at Berkeley (i.e., Data Science Education Program Student Teams/Peer Advising), Virginia Tech's DataBridge, or Emory's Center for Digital Scholarship which employs more than 30 graduate students. During the summer, there may be opportunity to shift the focus of these efforts to supporting students traditionally underrepresented in STEM who need to catch up on prerequisites (CALC 1, 2, 3, and linear algebra- likely to be required for the data science major). • Seek a dual appointment (co-funded position) for a data science librarian (Libraries and SILS).

Partnerships with Health Affairs (Curriculum and Research Integration): Develop partnerships to ensure data literacy skills are integrated at the point of need.
Potential Outcomes
<ul style="list-style-type: none"> • Review SPH & Pharmacy curricula maps to identify where data science is being taught and meet with curriculum committees and/or specific faculty to identify DS workshops and/or modules that can be integrated to support students. • Meet with Research Deans to identify data science needs related to research inquiry (e.g., use of tools to support systematic reviews, knowledge management, etc.) and offer opportunities for data science consultations. • Work with Global Women's Health (GWH) team to identify ways to integrate data expertise into the SOM/SPH study which will include development of algorithms, systematic reviews, and metadata application. • Provide expert searching and data analysis in support of research projects, systematic reviews, bibliometrics, and patient care.
Addressing Data Science with Library Collections (Research Integration): Provide purchased and curated discipline-specific data sets to support a wide variety of researchers.
Potential Outcomes
<ul style="list-style-type: none"> • 'Ready-to-use' data sets available for researchers, with documentation for each data set to orient researchers for their use, and learning modules featuring our collections as data (CAD). • Data purchasing program for Carolina researchers. • Partner with others on campus to build data literacy modules for the data literacy component of Triple I courses. • Establish a formal program for purchasing data sets including a method to solicit requests from researchers using the existing workflow for purchasing data sets. Begin the ordering and license negotiation process early in the fiscal year. • Create collections as data (CAD) for humanists. Advertise the data broadly and create learning modules featuring our CAD for courses. • Work with Research Computing and the rest of Information Technology Services (ITS) to make CAD "compute proximate" in the Cloud with appropriate identity management in place.

Coordinate Digital Humanities Initiatives (Research Integration): Center the Libraries as a unifying force for the many disparate digital humanities initiatives, experiments, and labs that pop up on campus.
Potential Outcomes
<ul style="list-style-type: none"> • Partner with Digital Humanities initiatives on campus (e.g., provide instruction to faculty groups to raise awareness of data science tools, methods, and resources on campus). This idea based on suggestion from UC Berkeley). • Framework to guide and encourage staff research projects. • Use the Digital South initiative as an incubator. • Provide space to potentially relocate one or more Digital Humanities Labs on campus. • Create analog to data (collections as data) workflow to facilitate the creation of data for humanists and formalize a service model for this work. This would build and expand on the collections as data work currently being done with the OneLibrary team spanning DRS, Special Collections, and L&IT. • Expand collaboration between Digital Research Services (DRS) and Software Development to provide a more robust set of skills on Digital Humanities research teams and grant funded projects. These teams may be composed entirely of Libraries staff or be composed of a mix of scholars from the libraries and elsewhere. This collaboration would expand on the expertise currently available from Digital Research Services to include deeper expertise in machine learning, database creation, and developer consulting/development time. • Establish an implementation team for the digital humanities recommendation, comprised of an expanded version of the Special Collections Digital Scholarship Working Group, to include additional staff members from software development, repository services, and DRS. The Digital South initiative should serve as an incubator for this recommendation, allowing librarians the space and time to engage in digital scholarship to support the initiative.
Collaborate with Scholars on Research Projects (Research Integration): Increase the Libraries' participation on research projects.
Potential Outcomes
<ul style="list-style-type: none"> • Expand our use of project charters. • Formalize a framework for including Libraries' staff as collaborators on research projects so we can respond to collaboration requests. • Increase awareness and understanding between the Libraries, Odum Institute, and RENCi to increase collaborations and facilitate referrals for consultation and research collaborations.

As DS @ Carolina emerges, we anticipate a **phased approach** in engaging stakeholders, where additional stakeholders are considered as new information is available from campus efforts. A preliminary list of stakeholders identified by the Committee as potential strategic partners for the Libraries around data science follows:

- Center for Faculty Excellence (CFE)
- College of Arts & Sciences
- Data Carpentries Program
- Digital Humanities (various groups across campus)
- Graduate and Professional Schools
- Institute of African American Research (IAAR)
- Information Technology Services (ITS)

- OASIS
- Odum Institute
- Office of Research
- Office of Undergraduate Research (OUR)
- Renaissance Computing Initiative (RENCI)
- School of Information and Library Science (SILS)

The Committee identified the following potential partnership opportunities around instruction, research, and other ways to collaborate.

Opportunities around instruction:

- Host datasets for use by courses to allow students to interact with data.
- Create data literacy curriculum for Triple I courses and supporting digital humanities projects.
- Increase focus on curriculum integration.
- Develop a data literacy curriculum.
- Develop workshops to help faculty integrate data into their courses.
- Expand teaching collaborations.
- Expand/formalize the use of Graduate Research Consultants for implementing data modules into courses.
- Introduce a data literacy component to the information literacy program for the Moore Undergraduate Research Apprentice Program (MURAP).
- Strengthen efforts to advance the use of data science methods in the humanities.
- Support faculty with integrating data modules into courses and ensure student success.
- Train humanities students interested in data science.

Opportunities to support research:

- Support research reproducibility and open science.
- Expand support for research reproducibility and open science.
- Collaborate on clinical research data projects.
- Develop toolkits (e.g., for grant support).
- Expand support for researchers to ensure research data comply with FAIR data principles (i.e., data are findable, accessible, interoperable, and reusable).
- Establish a data purchasing program for Carolina researchers.
- Investigate additional research collaboration opportunities.
- Promote research computing and data storage resources to facilitate researcher use.
- Advance the use of primary research using advanced text mining, database construction, and government data.
- Support graduate students throughout the research cycle.
- Train graduate students in research reproducibility, research data management, and open science.
- Expand efforts around research analytics, ORCID, and data management.

Other opportunities to form partnerships:

- Expand graduate research assistantships and CALA programs to support data science.
- Partner on efforts to promote diversity in data science (listed as a priority in their strategic plan).
- Partner on training opportunities.
- Reskill University Libraries' staff.
- Work together to develop a Faculty Learning Community for Data Literacy.
- Provide more joint consultations and collaborate on research teams.
- Increase awareness among campus partners to foster collaboration and facilitate referrals.
- Establish joint appointments.
- Partner and collaborate to make well-structured, large library data sets available to researchers.
- Partner to make collections as data (CAD) available for humanists.

As DS @ Carolina matures, it will be important for the Libraries to remain agile and phase in additional partners, understanding that the role we have with these partners may differ in terms of our level of integration. All partners selected should be cultivated with specific goals in mind and a clear idea as to what extent we wish to sustain, grow, or modify our level of partnership. When engaging these partnerships, the aim should be to raise awareness of our abilities and skills around data science, maintain communication, and find opportunities to increase our efficiency. The Committee developed a survey for assessing partner and stakeholder needs around data science ([Appendix D](#)) that can be used as the Libraries progress with the DS framework. Further, a Stakeholder Matrix ([Appendix E](#)) lays out needs by stakeholder type and level of expertise that could be utilized when determining in what ways the Libraries can collaborate.

The following is not an exhaustive list, but are some examples to consider:

- Carolina Population Center
- School of Media and Journalism
- Department of Philosophy
- Law School
- Center for Information Technology and Public Life (CITAP)
- NC Translational and Clinical Sciences Institute (NC TraCS)
- UNC Medical Center
- Lineberger Cancer Center

R4.B Communication with Partners



Establish Data Science Concierge Roles: Individuals should be designated as data science concierges to formalize collaboration and communication channels and ensure Libraries' are kept informed regarding data science activities around the libraries and more broadly on campus. Similar to how liaisons are embedded or have primary areas of connection, the data science concierges could help facilitate communication between Libraries data science efforts and partner needs and opportunities.



Formalize Mechanisms for Collaboration: Formalize use of project-based MOUs or other mechanisms to set expectations and timelines with campus partners. If applicable, individuals in

Data Science Concierge roles may be responsible for overseeing MOUs, including maintenance and process.

R4.C Joint or Co-Funded Positions

2

Joint-Funded Positions with Campus Partners: Use existing models of joint positions (e.g., SILS-CHIP/HSL appointment) to form new opportunities for co-funded positions. Expand and improve the existing models of adjunct professors to support departments/schools in their effort to teach specialized data science skills. Library staff teach courses regularly (e.g., Department of Geography, SILS) but incentives could be improved to expand these efforts.

Currently, the stipend to teach is low, time investment is high, and teaching must be done outside of work. Data science librarians could be beneficial as SILS and other units adapt to increased demand for enrollments in data science courses. Data Science RAs or GAs could assist with these efforts. Potential partners include departments in the College of Arts & Sciences, new Data Science Initiative/School, SILS, Odum Institute, or School of Medicine.

Recommendation 5: Creating and Expanding Services

With increased use of data science techniques, tools, and approaches to instruction and research, it is necessary for the Libraries to build capacity for select services and consider creation of new services. Service creation should be informed by a formal assessment with campus stakeholders.

New & Expanded Services Goals

- Meet Carolina's needs around data science.
- Cultivate data science partnerships.

R5.A Services to Expand

The University Libraries currently offers many data related services that are generally at capacity. Further, there is low redundancy in staff skills, making the services vulnerable to staff turnover. Reskilling and adding more staff is necessary to provide stability and additional capacity. To meet expected demand around these services, significant investment and cooperation with campus partners will be needed.

1 2

Expanding Services: The following services are currently offered by University Libraries and we anticipate that DS @ Carolina will increase demand in all instances. **Notably, we expect a significant increase in demand around predictive analytics using AI and machine learning, text analytics, data mining (particularly on library resources), and providing datasets for humanists from digitized primary source materials.**

- **Data Sourcing and Acquisition:** Locate and acquire appropriate data from external sources to help refine or answer research questions. This includes licensing, purchasing, describing, and processing data sets to make them available to researchers.
- **Data Creation:** Creation of data sets from a variety of sources including web scraping, data gathering via APIs, and structured data derived from digitized primary source materials.
- **Data Cleaning and Preparation:** Conduct all necessary transformations including merging, reshaping, or other reformatting to make analysis possible.

- **Analysis (AI and Machine Learning for Predictive Analytics):** Use machine learning techniques to predict outcomes/events, classify data, and identify data/literature sources (e.g., for a systematic review).
- **Analysis (GIS):** Use specialized tools and methods to work with geospatially referenced data (e.g., geovisualization, spatial analysis, and spatial statistics).
- **Analysis (Impact Measurement and Visualization):** Discover scope and pattern of research collaborations; measure and assess research impact by discipline area; communicate research impact to audiences such as funders or promotion and tenure committees.
- **Analysis (Inference Statistics):** Use of statistical methods to infer characteristics of a population or process.
- **Analysis (Text Analytics):** Use specialized tools and methods to derive meaningful information from unstructured text.
- **Analysis (Network Analysis):** Identify, measure, and visually represent relationship between groups of entities (people, terms, objects, etc.)
- **Visualization and Other Data Presentation:** Discover data insights and communicate findings through techniques that employ our innate ability to distinguish visual patterns in our environment.
- **Data Preservation and Archiving:** Using institutional repositories and other data storage systems to arrange, describe, and protect the provenance of data while preserving its integrity and making it available for reuse.
- **Data Management:** Organize and manage data during research; data documentation (e.g., metadata, file formats, naming conventions, file organization); version code or files (e.g., git).
- **Data Science Instruction:** Use of tools and methods (e.g., R, Python, visualization, GIS/mapping). Integration of data modules into existing courses.
- **Open Science:** Improve research by making processes and products open and accessible.
- **FAIR Data Principles:** Ensure research data are findable, accessible, interoperable, and reusable.

R5.A New Services to Create

As the Libraries grow to support data science, additional services will be necessary to support computing, open science/research reproducibility, and other types of analysis (e.g., image analysis, Genome-wide association studies (GWAS), meta-analyses, and music). Further, the Libraries will need to establish a data acquisition strategy and increase instruction around data literacy and data ethics. These services are currently not offered or only offered at a minimal level. To meet expected demand around these services, significant investment and partnership with campus partners will be needed.



Computing Support for Data Science Research: In partnership with Information Technology Services (ITS), University Libraries should explore services around computing, including research computing infrastructure such as specialized platforms to support processor-, memory-, or storage-intensive computing (high performance computing, parallel/distributed processing); data processing pipelines; and server operations/management.

1 2

Reproducibility: Provide support around documenting code with metadata to enable scientific replication or reproduction. Tools and methods for sharing code alongside data, with necessary information about compute environment, codebooks, etc.

1 2

Instruction Around Data Ethics: Instruction efforts should be expanded to support data literacy more broadly to increase capacity for integrating data modules into existing courses. Increased attention on instruction around ethical uses of data should be added around the following topics:

- Human Subjects
- Privacy/Confidentiality
 - Biometrics as personally identifiable information, HIPAA, etc.
- Algorithmic Bias
- Copyright (e.g., legality of web scraping)
- Social Impact of Data

Library values including diversity, equity, inclusion, and accessibility should be central to our approach for instruction in this area. Potential partners for developing instruction around data ethics include the Law School, Center for Information Technology and Public Life (CITAP), Hussman School of Media & Journalism, and Department of Philosophy.

1 2

Establish a Data Acquisition Strategy: Identify what role (and budget) the Libraries will have in acquiring data sets for researchers and teaching. Once established, augment and highlight existing efforts to purchase data for researchers – acquire, host, prepare, manage, and maintain specialized research data sets and ensure that the Libraries are involved in using data sets via curricular and research areas. The Committee recommends reviewing models such as the University of Michigan Data Acquisition for Data Science ([DADS](#)) program and University of Rochester River Campus Libraries [Datasets and Data Purchase program](#) as opportunities for further investigation.

1 2

Other Services to Consider Adding: The Libraries should explore additional services to offer including:

- Support for meta-analysis
- Image Analysis
- Data Curation
- Consolidated / Automated Referral System for common needs
- Biomedical Data Science
- Fast Healthcare Interoperability Resources (FHIR) standard for clinical data
- Citizen science initiative
- Biomedical-specific network analysis / visualization tools (e.g., Cytoscape, etc.)
- Gene expression analysis tools
- Open Science Framework
- Data support services directory

Recommendation 6: Space & Infrastructure

Much of the high-performance computing infrastructure required by data scientists is extraordinarily complex and expensive. The Libraries' Technical Infrastructure Review Team should focus on adding value to existing infrastructure in ways that leverage our core strengths.

Library spaces have been identified as a way to encourage use of services and facilitate collaboration. Our spaces will need to be examined to determine what we need to create, expand, or modify to support the initiative.

Space & Infrastructure Goals

- Cultivate community, interdisciplinarity, and catalyze new partnerships around data sciences.
- Establish a group to review the Libraries' current technical infrastructure.
- Create space use policies.

R6.A Library Space and DS @ Carolina

2

Create Formal Space Use Policy. A policy should be developed for library users who are unaffiliated with the data science program to use the spaces and infrastructure that we develop for this initiative. These policies should be as permissive as possible.

R6.B Technical Infrastructure

2

Technical Infrastructure Review Team. A team should be created to review the Libraries' current technical infrastructure for strengths and weaknesses related to implementation goals. This team would identify potential campus and external partners and define services related to our technical infrastructure and developer services.

2

Seek Infrastructure Partnerships. The Libraries' Technical Infrastructure Review Team should pursue partnerships around aggregation of related infrastructure providers, licensing, training, documentation, preservation, and the development of tools and practices to better integrate commercial infrastructure in the research lifecycle.

The Libraries should explore potential partnerships with campus Information Technology Services (ITS) units, including the Research Computing unit, to develop and manage service layers for campus research infrastructure.

2 3

Technical Consultation Services Program. The Libraries should develop a business model for providing specialized technical services based on internally-focused staff expertise, such as software development, digital preservation, digitization, and data management.

3

Tiered use policy for Libraries' technical infrastructure. To support the development of expanded and new services for data science, the Libraries will need to offer expanded and new technological services internally to library staff and possibly to the campus community. This policy should account for what technical infrastructure is available:

- for all UNC affiliates.
- for all UNC affiliates for a fee (see cost recovery recommendation).
- for library staff or projects in which library staff are documented partners.

The policy will also need to account for long-term management of projects that use Libraries' infrastructure. The policy should outline standard timelines for maintenance and project sunseting as well as project charter creation. Every project should include an agreement that details the agreed upon maintenance plan.

3

Business Plan to Manage Infrastructure Costs. To enable widespread use of our research infrastructure, the Libraries must develop methods to control costs for infrastructure that grow in proportion to use.

Appendix A: Executive Summary of Institutional Interviews

Over several months in Fall 2019, a subset of the Library Data Science Committee conducted interviews with library counterparts at several exemplar institutions, including MIT, NYU, Brown, University of Wisconsin Madison, UVA, and UC Berkeley. A high-level summary of these conversations is found below. (See [Appendix B](#) for guiding questions used during interviews.)

MOUs: The individuals we spoke to were not aware of MOUs or other tools being used to formalize project-based partnerships or joint appointments. Only one instance of a joint appointment was noted.

Staff Dedicated to Data Science: The number of staff dedicated to data science at each library ranged from 4-8 individuals.

- Several institutions indicated they are hiring or will be soon including non-MLS positions.

Partnerships: At the institutions we spoke to, strong partnerships exist between the library and:

- Clinical & Translational Sciences Institute (CTSI)
- Central IT data science group
- Research computing
- Graduate education
- Vice Chancellor for Research and Graduate Education
- Office of General Counsel (i.e., looking at institutional exposure and risk)
- Office of Sponsored Programs

Evaluating Data Science Services: Most individuals we spoke to indicated they would like to do more in terms of evaluating their data science services including evaluating longer term impacts.

- One institution noted that half of their research consultations have a data science focus and their data science course offerings fill up within 24 hours.
- Google analytics is used to track statistics where applicable.
- Annual surveys and qualitative assessments are used to evaluate services.
- LibAnalytics is no longer being used by one institution; they are currently considering new methods to evaluate impact.

Types of Services Offered: Services varied by institution. The following focus areas were identified during interviews:

- Partnering within the research lifecycle versus services that support the research lifecycle. For example, development of data management plans is not prioritized.
- Participation in the Carpentries community.
- Preservation, data management, ethics of data use, and data management program reviews.
- Digital scholarship workshops and 'toolkits'.
- GIS bootcamp.
- Web scraping and text analysis.

Effect of Data Science Programs on Library: Generally, libraries reported that data science initiatives at their campus have resulted in increased need for library data science services around:

- Big data
- Data curation
- Access, storage, backup, and management of large datasets
- Advanced text analysis
- Collaboration around data science instruction
- Workshops, tutorials, and consultations (A recent data-related workshop at one institution drew 80 attendees; this high turnout and interest has been sustained.)

Space and Infrastructure: Unsurprisingly, there was variation among the institutions in how the libraries provide space for data science initiatives. None of the individuals we spoke to indicated that they had exactly what they needed or wanted in terms of space for data science activities. Examples of library spaces that support data science include:

- Digital studio and collaboration areas with video wall; some computers with special software.
- “Research Commons” in the library that offers data science and other services; reservable, medium-speed computers.
- Data science service desk located in the Health Sciences Library that is in a great central location on campus.

Other Notable Information

The data science program at one institution is collaborating with their library to develop a Fellows program. Fellows will earn a **data science certificate**.

Multiple institutions use **Jupyter notebooks** for undergraduate courses.

Some **data sets** are bought by the library; some libraries create data sets; one institution indicated they are working with Web of Science to get access to raw data.

MIT has invested in and prioritized their liaisons to acquire data skills through:

- Providing **institutional membership with open science network**.
- Establishing **time-limited, team-based projects related to data**, including: data visualization, text and data mining, social justice issues related to data management).
 - Teams collectively identified learning projects through voting. Groups of 2-5 people devote 18 hours over 2-month period.
 - Goal is to position liaisons to take on a greater role with data services. Team-based learning has been effective.
 - Program concluded with an opportunity for participants to share their results with all Library staff.
 - Examples included:
 - Digging into federal health statistics to answer specific questions.
 - Generating maps.
 - Research to identify researchers using machine learning outside of the computer science program (e.g., ML in chemistry department).
- Providing **space and mandate** for staff to experiment with data and new skills.
- Focusing **learning directly on service areas** (versus learning for learning sake).
- **Removing boundaries** between data work and traditional liaison roles.

Appendix B: Interview Questions for Library Counterparts at Exemplar Institutions

Purpose: Understand how exemplar institutions have partnered with their data science programs.

Audience: Library Counterparts at nine exemplar institutions (see [Appendix F – Environmental Scan](#))

Method: Interviews via teleconference

The following questions were posed to one or more library champions at each institution for consideration.

Interview Questions
<p>1. What partnerships exist between the Library and other units on campus around data science? Follow-up where applicable: How did you go about cultivating the partnerships? Do you have any formal structures or frameworks that you use, such as joint appointments or MOUs? Follow-up if answer is no: What challenges have prevented this? (capacity, etc.)</p>
<p>2. Have you noticed any changes in the demand for library services due to the data science program/initiative at your institution? Follow up: If the library can't provide a data science service that is requested, who do you refer people to?</p>
<p>3. How did you determine the ways in which the library could be integrated in your institutions data science program (e.g., needs assessment, focus groups, surveys, data gathering, interviews/conversations, etc.)? Follow-up where applicable: Would you be willing to share some of your findings?</p>
<p>4. How many of your library's staff are affiliated with services that support data science work?</p>
<p>5. Did partnering with data science efforts require structural changes to your library org chart? If so, what new positions were created? How were existing job descriptions rewritten? Follow-up where applicable: Did departments or specific staff stop doing some types of work or services to make room for data science? If so, how did you communicate and manage that change?</p>
<p>6. What kinds of skills were necessary for your staff to have to meet the needs of data science efforts at your institution?</p>
<p>7. What methods does your library use to evaluate your partnership with data science efforts?</p>
<p>8. How would you describe your institution's readiness for change in this area when you started the program?</p>
<p>9. Did you have to rebrand existing services? How did staff and community/users react to the change?</p>
<p>10. Did data science efforts on campus require the library to change (or add onto) its technical infrastructure? Follow-up where applicable: Describe the process for this change.</p>
<p>11. Did you establish any physical spaces dedicated specifically to supporting this initiative? Did you partner with any other organizations to design or manage these spaces? Follow-up where applicable: In terms of spaces and services, have you found any areas of misalignment? between what you planned and actual use?</p>

Appendix C: Skills Matrix Survey

The survey below is an example of how to evaluate existing experience and proficiency in skills associated with the practice of data science. The directions and survey were designed for response from department heads. Sample output as a heatmap follows the survey and can be used to assess gaps in expertise quickly.

Directions

Complete the skills survey below for staff in your department. Refer to the explanations in this table when completing the survey.

Category	Explanation
Some Experience	Staff with <i>some experience</i> may have either (1) deep expertise in a very limited application of a certain topic or software or (2) has limited or intermittent general experience and the ability to tasks, with occasional help.
Expertise	Staff with <i>expertise</i> have significant experience with a method or software and the ability to quickly learn new tasks or help others troubleshoot.
Comments	Provide comments or additional information as desired. You may wish to note if staff with experience or expertise are primarily public-facing or work on internal library projects only.
Total	Indicate the total number of <i>individuals</i> with some experience or expertise in this cell. It is unlikely that the total number of individuals will be the total of this column (i.e., in many cases one individual will have experience in several areas or software tools).
Other	May include:
	3D Printing
	ArcGIS or QGIS
	Gephi
	GitHub
	Google Earth
	Java script
	Jupyter Notebooks
	Linux: shell scripting/bash
	Nvivo
	Solr
	Stata/SPSS/SAS
	Timeline JS
	VosViewer
	Wordpress
	Other automation tools (specify)

Survey

Skill or Software	Description	Number of University Libraries Staff with <i>some experience or exposure</i> in this area	Number of University Libraries Staff with <i>expertise</i> in this area	Comments
Data Sourcing	Staff who can find and compile appropriate data from external sources to help refine or answer the research question.			
Data Cleaning and Preparation	Staff who can conduct all necessary transformations including merging, reshaping, or other reformatting to make analysis possible.			
Analysis: AI and Machine Learning for Predictive Analytics	Staff who can use machine learning approaches to predict outcomes/events, classify data, identify data/literature sources (e.g., for a systematic review).			
Analysis: GIS	Staff who use specialized tools and methods to work with geospatially referenced data (e.g., coordinate systems, shapefiles and geodatabases, spatial statistics).			
Analysis: Impact Measurement and Visualization	Staff who can examine the scope and pattern of research collaborations; measure and assess research impact by discipline area; communicate research impact to audiences such as funders or promotion and tenure committees.			
Analysis: Inference Statistics	Staff who can use statistical methods to infer characteristics of a population or process.			
Analysis: Text Analytics	Staff who use specialized tools and methods to derive meaningful information from unstructured text.			
Analysis: Other Cases and Data Types	Staff with experience or expertise in other data types or analysis that requires specialized software or methods (e.g., image analysis,			


Skill or Software	Description	Number of University Libraries Staff with <i>some experience or exposure</i> in this area	Number of University Libraries Staff with <i>expertise</i> in this area	Comments
	Genome -wide association studies (GWAS), meta-analyses, music).			
Analysis: Network Analysis	Staff who can identify, measure, and visually represent relationships between groups of entities (people, terms, objects, etc.)			
Visualization and Other Data Presentation	Staff who discover data insights and communicate findings through techniques that employ our innate ability to distinguish visual patterns in our environment.			
Data Preservation and Archiving	Staff who can assist with using institutional repositories and other data storage systems to arrange, describe, and protect the provenance of data while preserving its integrity and making it available for reuse.			
Reproducibility	Staff who can document code with metadata to enable scientific replication or reproduction. Tools and methods for sharing code alongside data, with necessary information about compute environment, codebooks, etc.			
Data Management	Staff who organize and manage data during research; data documentation (e.g., metadata, file format, naming conventions, file organization); version code or files (e.g., git).			
Computing	Staff with experience or expertise around research computing infrastructure including specialized platforms to support processor-, memory-, or storage-intensive computing (high performance computing, parallel/distributed processing); data processing pipelines; server operations/management			

Skill or Software	Description	Number of University Libraries Staff with <i>some experience or exposure</i> in this area	Number of University Libraries Staff with <i>expertise</i> in this area	Comments
Integrating Data Literacy into Curricula	For example: Identifying learning objectives; creating a rubric for evaluation; updating syllabi; preparing instructional materials; and ethics, including: privacy/confidentiality; algorithm bias; copyright (e.g., legality of web scraping); social impact of data			
R	Open-source programming software emphasizing statistical analysis			
Python	Open-source programming software			
SQL	Relational database management system			
APIs	Application programming interface			
Tableau	Interactive data visualization software			
Advanced Excel	Examples: formulas; pivot tables, power query, etc.			
Other	Includes: 3D Design; GitHub; JavaScript; Library Carpentries Instructor; Linux: shell scripting/bash; MS Access; Oxygen; SharePoint; SPSS; VosViewer; Web management; Wordpress; XSLT and Xquery; Library Carpentries			
Total	Indicate the total number of <i>individuals</i> in this column (i.e., total number of individuals will likely be less than the total of the column as one individual may have experience or expertise with multiple skills or software).			

Sample Output

Heat map with sample output for survey to assess skills of Library staff. Cells shaded green indicate areas with significant experience or expertise and cells shaded red indicate areas where gaps exist.

Note: What constitutes high capacity will vary by institution; the number of individuals with experience or expertise in each area may also be added to heat map.

		
High Capacity		Low Capacity
Skill or Software	Number of University Libraries Staff with <i>some experience or exposure</i> in this area	Number of University Libraries Staff with <i>expertise</i> in this area
Data Sourcing: Compile appropriate data from external sources to help refine or answer the research question.		
Data Cleaning and Preparation: Conduct all necessary transformations including merging, reshaping, or other reformatting to make analysis possible.		
Analysis (AI and Machine Learning for Predictive Analytics): Use machine learning approaches to predict outcomes/events, classify data, identify data/literature sources (e.g., for a systematic review).		
Analysis (Network Analysis): Identify, measure, and visually represent relationships between groups of entities (people, terms, objects, etc.)		
Visualization and Other Data Presentation: Discover data insights and communicate findings through techniques that employ our innate ability to distinguish visual patterns in our environment.		
Data Preservation and Archiving: Assist with using institutional repositories and other data storage systems to arrange, describe, and protect the provenance of data while preserving its integrity and making it available for reuse.		
Reproducibility: Document code with metadata to enable scientific replication or reproduction. Tools and methods for sharing code alongside data, with necessary information about compute environment, codebooks, etc.		
Data Management: Organize and manage data during research; data documentation (e.g., metadata, file format, naming conventions, file organization); version code or files (e.g., git).		
Other (e.g., 3D Design; GitHub; JavaScript; Library Carpentries Instructor; Linux: shell scripting/bash; MS Access; Oxygen; SharePoint; SPSS; VosViewer; Web management; Wordpress; XSLT and Xquery; Library Carpentries)		

Appendix D: Survey for UNC Partners and Stakeholders

Purpose: Understand UNC's research and curricular needs around data science.

Audience: Campus partners & stakeholders.

Method: E-mail invite to Qualtrics survey for approximately 15 key UNC stakeholders/partners.

Question 1: Do you have current or anticipated needs around data science for research?

If yes → Answer Questions 3-5 and Questions 7-8

Question 2:

Do you have current or future needs around data science for instruction or curricula?

If yes → Answer Questions 6-8

If Yes to Question 1 (research focus)→

Question 3: What are your data science needs around research?

Question 4: Where do you go on or off campus for these services?

Question 5: To what extent are the following services important for your current and/or anticipated research needs?

Needs Around Research	Not important	Moderately Important	Extremely Important	NA	Comment
Data Sourcing <i>Find and compile appropriate data from external sources to help refine or answer the research question.</i>					
Data Cleaning and Preparation <i>Conduct all necessary transformations including merging, or other reformatting to make analysis possible.</i>					
Analysis: AI and Machine Learning for Predictive Analytics <i>Use machine learning techniques to predict outcomes/events, classify data, identify data/literature sources (e.g., for a systematic review).</i>					
Analysis: GIS <i>Use specialized tools and methods to work with geospatially referenced data (e.g., coordinate systems, shapefiles and geodatabases, spatial statistics).</i>					
Analysis: Impact Measurement and Visualization <i>Discover scope and pattern of research collaborations; measure and assess research impact by discipline area; communicate research impact to audiences such as funders or promotion and tenure committees.</i>					
Analysis: Inference Statistics					

Needs Around Research	Not important	Moderately Important	Extremely Important	NA	Comment
<i>Use of statistical methods to infer characteristics of a population or process.</i>					
Analysis: Text Analytics <i>Use specialized tools and methods to derive meaningful information from sources of unstructured text.</i>					
Analysis: Other Cases and Data Types <i>Any other data type or analysis that requires specialized software or methods (e.g., image analysis, Genome -wide association studies (GWAS), meta-analyses, music).</i>					
Analysis: Network Analysis <i>Identify, measure, and visually represent relationship between groups of entities (people, terms, objects, etc.)</i>					
Visualization and Other Data Presentation <i>Discover data insights and communicate findings through techniques that employ our innate ability to distinguish visual patterns in our environment.</i>					
Data Preservation and Archiving <i>Using institutional repositories and other data storage systems to arrange, describe, and protect the provenance of data while preserving its integrity and making it available for reuse.</i>					
Reproducibility <i>Document code with metadata to enable scientific replication or reproduction. Tools and methods for sharing code alongside data, with necessary information about compute environment, codebooks, etc.</i>					
Data Management <i>Organize and manage data during research; data documentation (e.g., metadata, file format, naming conventions, file organization); version code or files (e.g., git).</i>					
Computing <i>Research computing infrastructure including specialized platforms to support processor-, memory-, or storage-intensive computing (high performance computing, parallel/distributed processing); data processing pipelines; server operations/management.</i>					

Needs Around Research	Not important	Moderately Important	Extremely Important	NA	Comment
Data Creation <i>Creation of data sets from a variety of sources including web scraping, data gathering via APIs, structured data derived from digitized primary source materials.</i>					
Data Acquisition <i>Licensing/purchasing, describing, and making accessible data sets from vendors.</i>					
Other (specify)					
Other (specify)					
Other (specify)					

If Yes to Question 2 (curricula focus) →

Question 6: To what extent are the following topics important to your current or anticipated needs around curriculum?

Needs Around Instruction	Not important	Moderately Important	Extremely Important	NA	Comment
Ethics: Human subjects					
Ethics: Privacy/Confidentiality					
Ethics: Algorithmic bias					
Ethics: Copyright (e.g., legality of web scraping)					
Ethics: Social impact of data					
Training on specific tools					
Training around information literacy					
Training on terminology (i.e., jargon busting)					
How to find data					

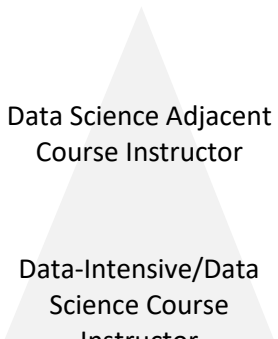
Needs Around Instruction	Not important	Moderately Important	Extremely Important	NA	Comment
How to share data					
Transparency & Reproducibility					
Communicating uncertainty					
Accurately communicating results					
Curriculum Integration: Identifying learning objectives around data literacy					
Curriculum Integration Creating a rubric for evaluation around data literacy					
Curriculum Integration: Updating syllabi around data literacy					
Curriculum Integration Preparing instructional materials for data literacy					
Curriculum Integration: Developing context <i>Teaching wise use of tools; matching assignments appropriately with level of students; defining scope; defining the research question; identifying questions that are answerable with existing data</i>					
Consultations with library staff around data literacy					
Other (specify)					
Other (specify)					
Other (specify)					
Other (specify)					
Other (specify)					

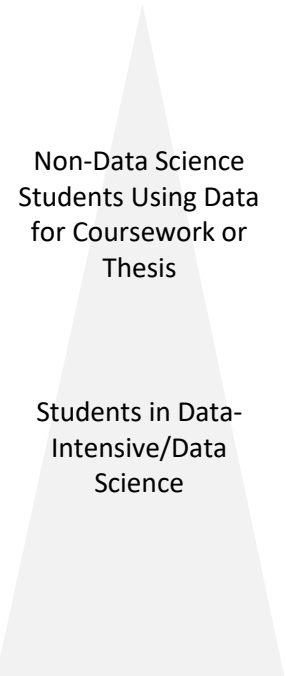
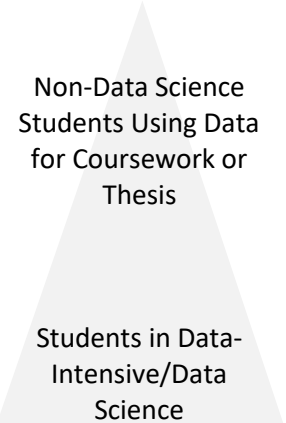
Question 7: Do you have additional comments you would like to share?

Question 8: Would you like to speak with us further? [add contact here].

Appendix E: Stakeholder Matrix

The stakeholder matrix summarizes types of UNC stakeholders and their potential needs by level of data science involvement. Needs would be met primarily through workshops, Open Labs, consultations, instruction, events (e.g., speaker series; Data Day event to share research), staff expertise, and library resources (e.g., purchased datasets).

Patron Type and Level of Data Science Involvement	Potential Needs
Faculty & Post-Doctorates (Researchers)	
 <p>Data Literate</p> <p>Data User</p> <p>Data-Intensive/Data Science Faculty</p>	<ul style="list-style-type: none"> • Awareness of resources & expertise available from the Libraries. • Identifying potential partners/collaborators (including colleagues from other disciplines and library partners). • Librarian consultations for projects requiring data expertise. • Physical space. • Support preparing Data Management Plans. • Support creating documentation. • Computing resources/infrastructure. • Purchased datasets. • Assistance overcoming methodological hurdles. • Assistance with publications. • Application of open research methods. • Long-term storage and preservation. • Handling sensitive data and de-identification.
Faculty (Instructors)	
 <p>Non-Data Focused Course Instructor</p> <p>Data Science Adjacent Course Instructor</p> <p>Data-Intensive/Data Science Course Instructor</p>	<ul style="list-style-type: none"> • Data literacy curriculum modules for different disciplines. • Course-specific instruction on data literacy. • Access to introductory resources providing data & visualizations. • Examples of classroom work. • Data, including digitized primary source materials. • Virtual space for course materials/data. • Pedagogical support. • Virtual sandbox for projects. • Specialized software for data manipulation and analysis. • Remedial support for students. • Problem sets. • Collaborative space and tools.

Patron Type and Level of Data Science Involvement	Potential Needs
Graduate Students	
 <p>Students in Non-Data Focused Disciplines</p> <p>Non-Data Science Students Using Data for Coursework or Thesis</p> <p>Students in Data-Intensive/Data Science</p>	<ul style="list-style-type: none"> • Instruction around data literacy, including data ethics. • Awareness and acquisition of data for all disciplines (including primary source materials). • Opportunities to learn about projects. • Communication skills for presenting research (e.g., through posters, presentations, visualizations). • Reputation management (e.g., via ORCID). • Small curricular modules for recitations. • Physical space. • Support locating and acquiring datasets. • Research assistance and project ideas. • Information about grant funding opportunities. • Information sharing to identify opportunities and increase visibility (e.g., for projects, employment). • Support preparing data management plans. • Support using open research methods. • Assistance overcoming methodological hurdles. • Connections to potential partners/collaborators (including interdisciplinary and library partners). • Handling sensitive data and de-identification. • Computing resources/infrastructure
Undergraduate Students	
 <p>Students in Non-Data Focused Disciplines</p> <p>Non-Data Science Students Using Data for Coursework or Thesis</p> <p>Students in Data-Intensive/Data Science</p>	<ul style="list-style-type: none"> • Communication skills for presenting research (e.g., through posters, presentations, visualizations). • Instruction around data literacy, including data ethics. • Opportunities to hear from fellow students. • Access to introductory resources providing data & visualizations. • Programming around real-world applications of data science, data ethics, and other topics. • Support locating datasets. • Virtual sandbox space. • Research support including advice on research plan. • Campus research computing resources. • Access to and support for specialized software for data manipulation and analysis. • Assistance overcoming methodological hurdles.

Appendix F: Environmental Scan

Executive Summary

Over 60 institutions were reviewed for their data science programs and partnerships with their university libraries and this summary focuses on nine universities, including:

- [Arizona State University](#) (Arizona State)
- [University of California at Berkeley](#) (Berkeley)
- [Brown University](#) (Brown)
- [University of Indiana](#) (Indiana)
- [Massachusetts Institute of Technology](#) (MIT)
- [New York University](#) (NYU)
- [University of Rhode Island](#) (Rhode Island)
- [University of Virginia](#) (UVA)
- [University of Wisconsin](#) (Wisconsin)

These universities were chosen because they are existing or emerging leaders around data science, have a special focus on artificial intelligence (AI), social science and the humanities, or data cataloging, or they have a stated connection to campus libraries or library school. Overall, no single university encapsulated all these aspects as part of their collaboration between their university library and a data science program.

Methods

The team started with a broad overview of over 60 institutions to get a sense of general trends and establish familiarity with baseline services. From this group, we prioritized for closer review 20 schools identified as notable in the Moore-Sloan Data Science Environments (MSDSE) report, including the 2018 “Academic Data Science Centers in the U.S.” report.

We also prioritized for closer review a list of schools included in a DataCure listserv discussion about universities with data science programs and librarians to support them, a small number of biomedical data science programs identified as noteworthy by Committee members with Health Sciences Libraries data services expertise, and programs highlighted in background readings provided to the Committee by Library leadership. From the initial survey, we selected nine schools and the thematic organization included in this document.

In our broad survey of institutions, we found that although there is a great deal of Data Science activity taking place across university campuses and libraries, there is room for greater integration of these programs.

To Identify Data Science Program-Library Partnerships:

- Checked Data Science programs’ webpages and looked for words such as “Partners” “Collaboration”, etc., when present. In rare cases, the Library was listed among the partners on these pages. Most of the time, these pages list units on campus such as other Centers/Institutes, High Performance Computing Centers, and industry partners.

- Searched DS programs' Events and News pages for Library-related posts. These searches turned up a few examples, primarily blog posts and announcements about collaborative projects and events.
- Searched in institutions' internal Data Science planning proposals/reports for Library connections.
- Overall, the MSDSE Project documentation provided the best sources, highlighting connections between programs and Libraries.

Results

Existing and Emerging Leaders

- MIT
- Berkeley
- UVA
- Indiana
- University of Wisconsin

Massachusetts Institute of Technology (MIT) – Institute for Data, Systems, and Society (IDSS)

Strengths

- Strongly aspires to build connections between computer and social sciences
- Leverages a research lab that is historically important and integral to the University
- Operates out of its own building on campus

Description of Data Science Program

While IDSS was formed to consolidate statistics-related programs across MIT, it does so while upholding a mission to connect efforts in Engineering with the Social Sciences and tackles complex challenges across multiple societal domains including finance, energy, infrastructure and health.

MIT's efforts to create a core area for statistics resulted in the **Statistics and Data Science Center (SDSC)** through which two of the IDSS degree programs are offered. In addition, IDSS is home to two research centers:

- **The Laboratory for Information and Decision Systems (LIDS)** which was established in 1940, making it MIT's oldest lab. LIDS now focuses on three main topics: systems and control, communications and networks, and inference and statistical data processing.
- **The Sociotechnical Systems Research Center (SSRC)**, whose mission is to "develop collaborative, holistic, systems-based approaches to complex sociotechnical challenges." ([website](#))

IDSS also provides a well-developed industry partnership program with multiple participation levels that allow companies to collaborate in both education and research.

Data Science Degrees Offered

- **MS in Technology and Policy (TPP)**
- **PhD in Social and Engineering Systems (SES)**
- **Interdisciplinary PhD in Statistics (IDPS)** through SDSC
- **Undergraduate Minor in Statistics and Data Science** through SDSC

Data Science Program Organizational and Physical Location

IDSS is housed in its own building which includes the LIDS and SDSC. It provides offices for faculty and students as well as spaces for collaboration.

Library Data Services, Activities, & Connections

The libraries at MIT provide a resource guide for IDSS that is managed by the Librarian for Electrical Engineering & Computer Science. No other specific connections between the library and the data science program are publicized. The Rotch Library houses a GIS & Data lab which also provides services for data management.

University of California, Berkeley – Division of Data Science and Information

Strengths

- Constitutes a broad network of connections across multiple governing bodies, academic programs and research initiatives
- Leverages and strengthens prior collaborative relationships among various groups throughout the university while continuing to build new ones
- Serves both graduates and undergraduates
- Works with libraries on campus to facilitate services

Description of Data Science Program

It is difficult to describe the entirety of the Data Science Division at Berkeley because its reach extends far beyond its academic programs both physically and influentially. Some notable parts of Berkeley's data science constellation include:

- **The Data Science Education Program**, serving undergraduate students as part of the College of Letters & Science
- **The School of Information**, serving graduate students
- **The Berkeley Institute for Data Science (BIDS)** which supports research, informal training and open source software development
- **The D-Lab**, a partner of the Data Science Division that provides consultations, training, working groups and meeting space
- **The Data Science Discovery Program** which allows undergraduates to participate in data research projects with graduate and post-doctoral students as well as community and entrepreneurial organizations
- **The Data Science Commons**, a new organizational structure still in development which will allow faculty from any part of the campus to propose the creation of a data science related program

Data Science Degrees Offered

- **Bachelor of Arts** with a major in data science from the College of Letters & Science
- **Master of Information and Data Science (MIDS)** from the School of Information

Data Science Program Organizational and Physical Location

It is unclear whether there is a specific building on the Berkeley campus that serves as headquarters for the Data Science Division; however, the integral parts of the program are in buildings near one another and central to campus. The College of Letters and Sciences, the School of Information, and the Doe Memorial Library (which houses BIDS) are all located next to each other.

Library Data Services, Activities, & Connections

The most notable connections between the University Libraries and the Data Science Division are BIDS, which lives at Doe Memorial Library, and the Data Peer Consulting program which operates out of the Moffet Library. Other sources for library data services include GIS support at the Earth Sciences & Map Library as well as consultants for data management and text analysis.

University of Virginia – School of Data Science (SDS)

The information in this section comes from UVA's [School of Data Science website](#) and accompanying the report: '[School of Data Science Phase II Faculty Senate Submission](#)'.

Strengths

- Integration between UVA Library and the School of Data Science (SDS) figured prominently in the SDS Phase 2 Proposal, which was unanimously approved by the Faculty Senate ¹
- Proposal identified UVA's/UVA Libraries' Digital Humanities (DH) strengths as a key interdisciplinary area for further exploration
- SDS aims to build and maintain the new School as an open scholarly ecosystem, citing the opportunity to establish UVA's leadership in Open Scholarship as an important motivating factor
- SDS plans to serve both undergraduates and graduates

Description of Data Science Program

In June 2019 the Virginia Board of Visitors approved a new School of Data Science (SDS) and it will focus on:

- Responsible data science
- Diversity, openness and transparency
- "Open scholarly ecosystem" for the public good: aiming to openly share policies, procedures and educational materials, lab materials, data, analytics published literature

SDS's focus on the public good will build on UVA's earlier Data Science Institute's (DSI) work in these areas, such as:

- Global Women in Data Science Conference regional ambassador
- Data for the Social Good service learning/community outreach initiative --- SDS faculty, staff, students and alumni developing tools for matching community non-profits with students and service-learning projects providing data analysis.

The ability to grant tenure for recruiting faculty belonging solely to the SDS was a key reason for establishing the new School. The SDS will also continue to have joint-appointment faculty, as well as Fellows and researchers from within and outside UVA.

Data Science Degrees Offered

- **Master of Science in Data Science**, residential and online
- **Dual-degree master's programs**: Business (MBA/MSDS), Medicine (MD/MSDS), and Nursing (PhD/MSDS).
- SDS plans to offer undergraduate, Ph.D., certificate and executive education programs.

¹ <https://news.virginia.edu/content/uva-board-approves-establishment-school-data-science>

Data Science Program Organizational and Physical Location

SDS is a new, independent School within UVA, but SDS's organizational design is meant to encourage integrations with other schools as much as possible:

- The school will operate satellites and centers embedded in other schools, instead of departments.
- Centers will be theme-based, covering areas of data analytics; data visualization and dissemination; democracy; deep learning; education; and ethics, policy and law. ²

UVA announced plans to build a new 70,000-square-foot academic building for the SDS, within a 14-acre parcel developed “around three interrelated nexuses – creativity, democracy and discovery (...). The corridor will ultimately house other departments and initiatives, visual and performing arts spaces, and a hotel and conference center.” ³

Library Data Services, Activities, Connections

UVA Library's [Research Data Services + Sciences](#) pages provide excellent, easily navigable documentation for the Library's presence across the scope of Research Data Services. In particular, [the FAQs page](#) is a comprehensive and accessible model for organizing a complex set of services.

Both the SDS website and the internal Phase II proposal provide examples of already-established functional connections between SDS and the Library, and a strong partnership mindset overall.

- [SDS blog post](#) about [The Open Data Lab](#), a SDS-Library partnership: "a data sharing network that will support research through its entire process — from the data collection stage through analysis and data use."

Library connections highlighted within the Phase II proposal's overall vision for integrating the SDS across all UVA schools:

- The **Library is included as a principle collaborator**, at the same level as planned SDS satellites, in the proposal's 'Example Collaborations (...)' table (Table 7, p. 45)
- **SDS-Library existing collaborations on training and research projects**: Scholars Lab, Scholia @ University of Virginia, and plans to expand UVA's presence in Wikimedia projects: Wikipedia - Trust and safety, and the Cochrane Wikipedia Partnership. (p.51)
- Digital Humanities and social sciences collaborative work with the Library are cited as areas to build capacity (p.61)
- SDS planned organizational structure **includes the Library within the Operational arm**: "Data and Information Resources (...) a special functional module to work with the library and IT services to cooperatively make available data, analytics, information etc. both needed by the SDS team (faculty, staff, students) and offered by the team." (p.64)

The Phase II proposal also highlights SDS's planned 'Open Scholarly Ecosystem' as an opportunity to establish UVA as a leading US and global academic institution in open scholarship. DSI will provide "a

² <https://news.virginia.edu/content/uva-faculty-senate-votes-establish-school-data-science>

³ <https://news.virginia.edu/content/uva-board-approves-establishment-school-data-science>

major economic driver for the Commonwealth of Virginia and beyond while at the same time providing accessible knowledge to a diverse audience.” (p. 74)

The proposal points to existing and planned Library partnership activities as key to supporting a complete open research lifecycle (all points below from p.76):

- UVA’s innovative library and association with Hathi Trust has already laid a foundation for UVA as a leader in open scholarship
- Open publication of research will be encouraged, supported by collaboration with UVA Library, and platforms such as Ubiquity or Coko
- Encouraging/providing open teaching materials will be supported by collaboration with UVA Library, for open e-book services for textbooks, other materials
- Deposit into the Libra IR (managed by UVA Library) will be encouraged, to preserve all completed SDS research

In terms of connections from the Library’s side, the Stat Lab’s [Related Resources](#) page includes the DSI, as well as many other University Resources/Communities.

University of Indiana – Data Science Program

The Data Science Program ([website](#)) is an interdisciplinary collaboration between four departments within the School of Informatics, Computing and Engineering (SICE) and the College of Arts & Science (Statistics)

Strengths

- Interdisciplinary at all levels
- Program covers BS, MS, PhD (minor) and certificate
- Serves both undergraduate and graduate students
- Located within the School of Informatics, Computing and Engineering (SICE)
- Data research happens at the Data to Insight Center (D2I) a collaboration between SICE, the University Libraries and the Pervasive Technology Institute

Description of Data Science Program

The data science program is interdisciplinary and collaborative across several departments in SICE, the College of Arts & Science, the O’Neill School of Public and Environmental Affairs, the Kelley School of Business and the School of Public Health.

Data Science Degrees Offered

- **BS in Data Science** from SCIE
- **MS in Data Science** from SCIE, residential and online; two learning tracks, either Applied Data Science or Computational and Analytical
- **PhD (minor) in Data Science** (12 credits) from SCIE, residential or online
- **Certificate in Data Science** (12 credits) –from SCIE, online

Data Science Program Organizational and Physical Location

The Data Science Program is an interdisciplinary collaboration between four departments (Computer Science, Informatics, Information & Library Science and Intelligent Systems Engineering) within SICE and the College of Arts & Science (Statistics). The Data Science program is housed within SICE.

Library Data Services, Activities Description

The services offered by the library Data Services unit ([website](#)) works with researchers at all stages of a project; collecting, processing, analyzing, publishing and preparing data for long term storage. This unit also provides GIS and Statistical Data Services.

- Collaborators include the Political Science Data Laboratory and Archive, the Karl F. Schuessler Institute for Social Research, and the Stat/Math Center of University Information Technology Services (UITS).

University of Wisconsin, Madison – The School of Computer, Data & Information Sciences (CDIS)

The School of Computer, Data & Information Sciences ([CDIS](#)) is the newest division in the College of Letters & Science and was launched September 2019. It is a collaboration between the departments of Computer Sciences, Statistics and the Information School (iSchool). There will be no change to the leadership or autonomous governance structures of the three units.

Strengths

- Interdisciplinary programs across campus; pan-campus including the sciences, social sciences and humanities
- Plans include serving both undergraduate and graduate students
- Focused on the needs of the Wisconsin including business, job creation, educating and serving the public, and outreach
- Research areas include AI/machine learning, social and ethical aspects of computing and data science, human-computer interaction (design) and cybersecurity
- CDIS will partner with American Family Insurance Data Science Institute ([website](#))

Description of Data Science Program

Currently, the data science program is in development:

- Plans for joint degrees, certificates and classes at the undergraduate and graduate levels.
- Initial proposals include an undergraduate program in data science and a master's in information studies

Data Science Degrees Offered

- **MLIS with a Data/Information Management & Analytics (DIA)** concentration at the iSchool
- **BS/BS, PMP, MS/PhD** in Computer Science
- **MSDS** Master's in Statistics with data science

Data Science Program Organizational and Physical Location

- Collaboration between the departments of Computer Sciences, Statistics and the Information School (iSchool). There will be no change to the leadership or autonomous governance structures of the three units.
- No physical building at this time.

Library Data Services, Activities, & Connections

The Research Data Services ([RDS](#)) is an interdisciplinary organization, which provides research data management practice across campus. It includes the libraries, the Division of Information Technology,

the Office of the Chief Information Officer, and the Institute on Aging. It provides researchers with the tools and resources that support their efforts to store, analyze, and share data.

There is a Data Science Hub, coordinated by RDS, which offers workshops in a variety of data science areas and in various locations across campus (based on information provided on the [website](#)).

Connections include:

- American Family Insurance Data Science Institute

Artificial Intelligence (AI)

University of Rhode Island Data Science Program and Artificial Intelligence Lab

Strengths

- The AI Lab at URI is a product of library staff working hand-in-hand with faculty in computer science and engineering.
- The Lab is engaged with both curriculum at the university and education programs for the public.

Description of Data Science Program

Data Science at Rhode Island has a higher concentration of support at the undergraduate level, offering both a BA and BS through the Department of Computer Science and Statistics. The department is connected to the Artificial Intelligence Lab through its faculty.

The AI Lab is the first of its kind to emerge from a university library and is open to both university members and the public. It provides access to a high-performance supercomputer, six laptop workstations, and a small but growing collection of IoT devices. Training at the AI lab is available through workshops but is also closely integrated with multiple university courses across various disciplines. In addition, the Lab facilitates summer camp programs for K-12 students in the community.

Data Science Degrees Offered

- **Bachelor of Science in Data Science**
- **Bachelor of Arts in Data Science**
- **Undergraduate Minor in Data Science**
- Related graduate degrees in areas such as computer science, statistics, cyber security and digital forensics

Data Science Program Organizational and Physical Location

The AI lab is housed in the Robert L. Carothers Library which is a short walk from the Computer Science department in Tyler Hall. The library is also home to a makerspace with 3-D printers, allowing users full access to equipment necessary for wearable technology and other device creation.

Library Data Services, Activities, & Connections

Multiple librarians are counted among the founding members of the AI lab and library staff are responsible for its management alongside faculty members in computer science and engineering. Outside of the Lab and the makerspace, the library offers a digital repository, but no specific data services.

Social Sciences / Humanities

- Brown

- MIT

Massachusetts Institute of Technology (MIT) – Institute for Data, Systems, and Society (IDSS)
[See above.](#)

Brown University - Data Science Initiative (DSI) at Brown University

Brown was not included in the Academic Data Science Centers Report; this assessment draws primarily from Brown’s DSI and Library web pages, and the University newspaper.

Strengths

- Visibility and outreach at many levels within DSI, such as browsable, accessible presentation of interdisciplinary Research Projects, outreach events with campus partners
- Effort to overcome silos: Library inclusion in shared Digital Teaching and Learning Resources
- DSI activities demonstrate clear alignment with mission, focus on applying data science to cultural and social spheres
- Serves both undergraduates and graduates

Description of Data Science Program

The DSI’s mission encompasses “both domain-driven and fundamental research in data science,” and prioritizes “the impact of the data revolution on culture, society, and social justice.”

DSI’s Research Grants program focuses on new initiatives and collaborations, particularly interdisciplinary work across disciplines or units, and encourages submissions for projects emphasizing the public good, in alignment with the DSI’s mission. DSI showcases a range of Research projects at a well-designed, browsable [Research Projects](#) page, linking to the Abstract, Project Leads, and Funding Source for each. Some relevant examples:

- [Social and Family History - Extraction, Representation, and Evaluation:](#)
 - leverage social, behavioral, and familial data from electronic health records to create rich longitudinal resources, informing understanding about various determinants of health
- [Computational Psychiatry: Combining Theory-Driven and Data-Driven Approaches to Understand Impulsivity](#)
 - Use neuroimaging with mathematical modeling to discover cognitive neuroscience mechanisms underlying impulsivity, as a substantial risk factor for aberrant behaviors
- [Predictive Healthcare Analytics](#)
 - build models of complex health phenomena from biological, clinical, and public health data, emphasizing pediatrics, psychiatry, emergency medicine, and critical care

DSI’s Events program is notable for collaborations across the University, and for addressing social justice issues and diversity/equity/inclusion. Recent topics in the [robust Events calendar](#) included

- a “hands-on open house featuring many of our Illustrating Mathematics program participants”
- a statistics seminar on “Useful models of mental and emotional functions for an increasingly detailed picture of the brain-body connection and its many roles in health and survival.”
- a training installment in the Research Integrity Series: The Role of the Scientist in Society”.

Other examples of outreach-oriented DSI events:

- [“Algorithmic Justice: Race, Bias and Big Data”](#) panel, cohosted by the University’s Center for the Study of Race and Ethnicity in America and the Data Science Initiative
- DSI participation in the Brown Graduate School’s [Academy in Context](#) seminar series - DSI Director presented on ‘Bias and Transparency in Contemporary Data Science’

Data Science Degrees Offered

- **Master’s Degree in Data Science** (Master of Science, ScM) - a twelve-month program (September-August), offered by DSI since the Fall of 2016
- **Doctoral certificate** for current Brown PhD students in other fields.
- **‘5th-Year Master's Degree’** - Brown undergraduates are eligible to apply to the MA program, allowing two credits to be substituted by undergraduate coursework.

Data Science Program Organizational and Physical Location

The DSI includes the departments of computer science, mathematics, applied math and biostatistics. DSI can grant degrees but not tenure; faculty have joint appointments.

Since 2018 the DSI has occupied space within a large office building that was renovated for the Robert J. and Nancy D. Carney Institute for Brain Science. Along with Carney Institute staff, and laboratories for departments pursuing computational neuroscience, human brain recording and decoding, and neuro-engineering, the Center for Computational Molecular Biology also shares space in this building.⁴

Library Data Services, Activities, and Connections

Brown’s Library data services include: data repository/preservation, DMP, visualization, support for teaching with digital methods, GIS and Storymap, Lab notebooks.

The Library’s Events page includes workshops and support for many specific tools for working with data. Recent data workshops covered: Visualization, Publishing, DMP Tool, Best Practices, LabArchives Notebooks, Finding Funding Opportunities, Software Carpentries, and Introduction to Topic Modeling for the Humanities.⁵

A particularly strong point is the Library’s Data Services inclusion in a well-organized [Digital Teaching and Learning Resources](#) (DTLR) site. DTLR gathers resources from organizational units across the University, including: Brown’s Sheridan Center for Teaching and Learning, The Library and the Library’s Center for Digital Scholarship, and CIS Academic Technology.⁶

- Shared Resources page makes a wide range of services that are usually scattered and siloed easy to find, and seems likely to increase visibility for Library data services

The DSI co-sponsors with the Brown Libraries and the Initiative for Computation in Brains and Minds a [Data Science, Computation & Visualization \(DSCoV\) workshop series](#): “introductions to basic data science and programming skills and tools, offered by and for Brown staff, faculty, and students (with occasional presenters from outside Brown).”

⁴ <https://www.brown.edu/carney/news/2019/01/18/carney-opens-new-home-innovation-and-impact-brain-science>

⁵ <http://brownlibrary.lwcal.com/#view/all>

⁶ <https://www.brown.edu/academics/digital-teaching-learning/about-digital-teaching-learning-brown>

DSCoV Workshop materials are collected in a growing, publicly available [Github repository](#). Examples of recent topics: Basics of Data Exploration and Visualization with Vega, Tidying, Transforming, and Visualizing Data in R, Deep learning in Kaggle Kernels.

Data Cataloging

New York University – Center for Data Science

Strengths

- CDS-Library partnership across many areas of common interest, especially strong within NYU Libraries' focus on reproducible research and data curation
- CDS work in software development, training and events, community outreach
- New DS3 service: CDS working with NYU Libraries and other partners to provide skilled labor to data science research projects
- Serves undergraduates, graduates, and non-degree professionals

Description of Data Science Program

NYU's Center for Data Science (CDS) was established in 2013. It was significantly expanded through participation in the Moore-Sloane Data Science Environments program, receiving 5 years of support, 2014-2019.

CDS cannot grant tenure. Faculty have joint appointments with "home" departments. CDS's faculty page includes 16 joint-appointed faculty, over 40 affiliated faculty, and a few each of Associated, Visiting, and Adjunct faculty.

Data Science Degrees Offered

- **Master of Science in Data Science** launched in 2013 and has grown quickly: "in 2017, there were 1,800 applicants for 150 slots." (ADSCUS p.31)
- **PhD Degree** program added in 2016
- interdisciplinary **Undergraduate Major and Minor Programs in Data Science**: "Both programs of study are open to all undergraduate students from across the humanities, social sciences, and sciences, and our courses are open to all students."
- **Non-degree program** for professionals

Data Science Program Organizational and Physical Location

CDS moved into a newly renovated space in 2016. It occupies two floors: one for "quiet work", one for events and collaborative projects. (ADSCUS p.31)

Library Data Services, Activities, and Connections

NYU's data services instruction guide provides an outstanding model:

- covers a wide range of software, processes, and types of data analysis
- organizes classes into suggested groupings and sequences to support various stages of the search cycle for working with data.

A [Data Science Classes Diagram](#) organizes classes into groupings based on the data research cycle: collect/find, clean & organize, explore, analyze, and share. These stages allow for further sub-groupings, such as Analysis courses to support different types of data: GIS, Qualitative, Quantitative, or HPC. The

data services diagram reveals a coherent instruction program through which researchers can navigate according to their specific needs, and build skills sequentially.

In 2019, NYU joined the [Data Curation Network](#), which provides data curation services and education for support building FAIR (findable, accessible, interoperable and reusable) research data collections in institutions' repositories.

The Libraries' Data Services and Digital Scholarship Services co-host the [annual DH + DATA Day](#) which includes a [mapping and visualization competition](#) voted on by all DH + Data Day participants.

The Academic Data Science Centers report outlines significant areas of partnership between CDS and the NYU Libraries (page 32):

- "In partnership with the NYU Libraries, CDS has also been actively involved in promoting reproducible science practices on campus by offering training sessions, consultations, and lectures on how to obtain and manage large datasets."
- CDS staff have contributed software development expertise and time to supporting open-source software projects for computational reproducibility, such as Reprozip and ReproServer, which allow researchers to package their data and computation software together for later execution/reproduction of results.

The ARL Report points to additional areas of collaboration (all points below from p.20):

- CDS worked closely with the libraries to design its new permanent space.
- The library's deep understanding of how people across the campus and across disciplines use space—for writing, reading, interacting with technology, and collaborating, and experience with modularity and flexibility in space design informed CDS's new environment.
- "the growth of data science throughout the university has influenced the library's collecting, such as purchasing more vendor produced data sets, responding to students' need for big data (for example, large social media feeds), and integrating APIs into their collection and discovery environment."

NYU Libraries, the CDS, and NYU IT Research Technology are piloting **Data Science & Software Services (DS3)** in connection with related campus units including High Performance Computing (HPC), and NYU IT Teaching and Learning with Technology (TLT). DS3 is a [new joint service](#) providing centralized access for faculty and research staff needing data scientists to work on grant funded projects that require expertise in data analytics, statistical methodology, and research-oriented software development.

DS3's stated goal is "to enhance the research capacity of NYU by providing highly skilled labor for funded projects and increasing the competitiveness of grant proposals."

Service areas draw on the NYU Libraries' and CDS's combined strengths (all points below from <https://cds.nyu.edu/ds3/>):

- Research methodologies in Data Science and Statistics
- Scoping development activities
- Experimental and research design
- Software development for research

- Data analytics for all data types and sizes

Stated Partnership with Campus Libraries or Library School

- Arizona State
- Indiana

University of Indiana – Data Science Program

[See above.](#)

Arizona State University

Arizona State University (ASU) offers data science related degrees in the departments with residential and online programs. Finding information about all degrees offered at ASU was difficult and we may not have found all of them.

ASU has a certificate in data science but does not have a specific data science program, but the library has a Data Science and Analytics Unit which has many partners across campus.

Strengths

- Data science related degrees through the business school
- Serves both graduates and undergraduates

Description of Data Science Program

Data science at ASU is part of existing school so there is not a separate school or department.

Data Science (related) Degrees Offered

- **Bachelor of Science** in Business Data Analytics at W.P. Carey School of Business, online
- **Master of Science in Business Analytics (MS-BS)** at W.P. Carey School of Business
- **Certificate in Data Science** from New College of Interdisciplinary Arts & Sciences ([website](#))

Data Science Program Organizational and Physical Location

- No building or location

Library Data Services, Activities, and Connections

- Data Science and Analytics Unit within the Hayden Library ([website](#))
- Connect experts to methods and technologies of data science

The library hosts a weekly “open lab”⁷ to introduce data science to students, staff and faculty and to work with groups on research projects that covers many subject areas. The library has many collaborators across the campus including but not limited to:

- Department of English
- Global Biosocial complexity Initiative
- Herberger Institute of e Design and the Arts
- School of Computing, Informatics and Decision Systems Engineering

⁷ <https://lib.asu.edu/data/open-lab>



Library Data Science Committee

Framework Recommendations

University Libraries
The University of North Carolina at Chapel Hill
March 2020

Committee Members

Nandita Mani
Michelle Cawley
Lorin Bruckner
Jason Casden
Adam Dodd
Amanda Henley
Matt Jansen
Jamie McGarty
Morgan McKeehan
Sarah Morris
Therese Triumph
Jessica Venlet
Joe Williams